

AMENDMENTS TO THE CLAIMS

A detailed listing of all claims that are, or were, in the present application, irrespective of whether the claim(s) remains under examination in the application are presented below. The claims are presented in ascending order and each includes one status identifier. Those claims not cancelled or withdrawn but amended by the current amendment utilize the following notations for amendment: 1. deleted matter is shown by strikethrough for six or more characters and double brackets for five or less characters; and 2. added matter is shown by underlining.

1. (Cancelled)
2. (Cancelled)
3. (Previously presented) The method of Claim 23, wherein said resource capability information comprises a resource capability metric for each of at least five resource parameters for each server, regarding the server's CPU availability, memory availability, connectivity to storage, connectivity to a proxy server, and connectivity to peer servers.
4. (Previously presented) The method of Claim 3, wherein said method further comprises the step of providing a dynamic, relational database and process of statistical inference for ascertaining expected demand patterns involving said at least one attribute.
5. (Original) The method of Claim 4, wherein the number of said expected demand patterns can dynamically increase or decrease.
6. (Original) The method of Claim 4, wherein a resource requirement metric for each of said at least five resource parameters is assigned to each said expected demand pattern, wherein each said collected request is identified with at least one said expected demand pattern, and wherein said resource requirement metrics assigned to said identifying expected demand pattern are further assigned to said collected request.

7. (Previously presented) The method of Claim 6, further comprising the steps of determining the metric distance between said resource requirement metrics and said resource capability metrics for at least one combination of said collected client request and server pairings and selecting a server for each said collected request so that the sum of said metric distances for said at least one combination of said pairings is minimized.

8. (Previously presented) The method of Claim 7, wherein an optimization paradigm is used to at least partially perform the step of selecting a server for each said collected request so that the sum of said metric distances for said at least one combination of said pairings is minimized.

9. (Canceled)

10. (Canceled)

11. (Canceled)

12. (Canceled)

13. (Canceled)

14. (Canceled)

15. (Canceled)

16. (Canceled)

17. (Canceled)

18. (Previously presented) The system of Claim 24, wherein said resource capability information comprises a resource capability metric for each of at least a plurality of resource parameters for each server selected from the set comprising: the server's CPU availability,

memory availability, connectivity to storage, connectivity to a proxy server, connectivity to peer servers, and connectivity to the Internet.

19. (Currently Amended) The system of Claim 24, wherein the relational database is dynamic and wherein the number of said ~~expected demand~~ patterns can dynamically increase or decrease.

20. (Currently Amended) The system of Claim 24, wherein ~~[[a]] the resource requirements metric-vector associated with one request among the client requests incoming within the predefined time interval is for each of said at least five resource parameters is assigned to each of said expected demand pattern, wherein each said collected request is correlated with at least one said expected demand pattern, and wherein said resource requirement metrics assigned to said correlated expected demand pattern are further assigned to all said collected client requests incoming within the predefined time interval.~~

21. (Currently Amended) The system of Claim ~~[[20]]~~ 24, wherein the optimization and allocation process for selecting a server resource for receiving each client request is configured so that the sum of all costs in the cost matrix is minimized for the process of distributing said collected requests further comprises determining the metric distance between said resource requirement metrics and said resource capability metrics for at least one combination of a pairing of said collected request and server and selecting a server for each said collected requests so that the sum of said metric distances for said at least one combination of said pairings is minimized client requests and server resources.

22. (Currently Amended) The system of Claim ~~[[21]]~~ 24, wherein a global optimization paradigm is used to minimize a total assignment cost for ~~said pairings of all of said servers and said requests~~ the said all combinations of client requests and server resources.

23. (Previously Presented) A method for distributing incoming client requests across multiple servers in a networked client-server computer environment comprising:

- (a) collecting client requests incoming within a predetermined time interval;
- (b) upon receipt of the at least two incoming client requests within the predetermined time interval, analyzing each of the client requests using categorical criteria and demographic criteria to extract attributes of the request;
- (c) classifying the client requests based on the extracted attributes by comparing each request with a pattern selected from a set of patterns in an adaptive request table to find a match-pattern that best matches the request;
- (d) using the match-pattern to associate a requirements vector with each request, the requirements vector being populated with at least five resource parameters that prescribe the expected resource requirements of the request;
- (e) capturing resource capability information for each server at least once during the predetermined time interval, each server being associated with a capability vector refreshed with the resource capability information;
- (f) following steps (d) and (e), for each client request and server pair, computing a vector space distance between the requirement vectors and capability vectors corresponding to the client request and the server respectively, the vector space distance being an update to an element in a cost matrix initialized at the start of the predetermined time interval;
- (g) at the completion of the time interval, distributing the client requests across multiple servers to minimize a cost metric associated with the cost matrix for all combinations of client requests and server resource capabilities; and
- (h) repeating steps (a) through (g) for subsequent ones of said time interval after

initializing the cost matrix.

24. (Currently Amended) A system for servicing multiple requests to be distributed across multiple servers in a networked client-server computer environment, the system comprising:

a request examiner routine for analyzing each of the client requests incoming within a predefined time interval, using categorical criteria and demographic criteria to extract attributes of the request;

a request table for collecting attribute information associated with each client request incoming within the predetermined time interval;

an adaptive request table populated with a set of patterns, each pattern associated with a generic request type that is most likely to be received by a proxy server, the adaptive request table suited for classifying client requests by comparing each request with a pattern selected from the set of patterns in the adaptive request table to find a match-pattern ~~an expected demand pattern~~ that best matches the client request resulting in a successful match, so that an expected demand on resources by the client request can be predicted using the ~~expected demand~~ match-pattern in the adaptive request table to associate a requirements vector with each request, the requirements vector being populated with at least five resource parameters that prescribe the expected demand on resources imposed by the request;

a relational database coupled with a process of statistical inference to facilitate the construction of an adaptive request table, the adaptive request table being updated by the process of statistical inference for each match-pattern resulting from a successful match of the client request with ~~the expected demand pattern from~~ the set of patterns;

a resource table for collecting resource capability information about each server at least once during the predetermined time interval wherein each server is associated with a capability vector refreshed with the resource capability information;

a cost matrix initialized at the start of the time interval, the cost matrix characterized by a row index associated with a request identifier and a column index associated with a resource identifier, each element in the cost matrix representing a vector space distance between the requirements vector corresponding to the client request and the capability vector corresponding to the server resource;

a distance vector computing process for computing the vector space distance between the requirements vector and the capability vector for each pairing of the client request and the server resource; and

an optimization and allocation process for selecting a server resource for receiving each client request so that a sum of all costs in  $[[a]]$  the cost matrix is minimized for all combinations of client requests and server resources.